|  |  |
| --- | --- |
| C:\Users\USER\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\616878FB.tmp | **Министерство науки и высшего образования Российской Федерации**  **Федеральное государственное бюджетное образовательное учреждение**  **высшего образования**  **«Московский государственный технический университет**  **имени Н.Э. Баумана**  **(национальный исследовательский университет)»**  **(МГТУ им. Н.Э. Баумана)** |

|  |  |
| --- | --- |
| ФАКУЛЬТЕТ | Информатика, искусственный интеллект и системы управления |
| КАФЕДРА | Системы обработки информации и управления |

**Отчет по лабораторной работе №2**

**«Обработка признаков (часть 2)»**

Студент группы ИУ5-25М

Денисов Алексей Сергеевич

Москва, 2023 г.

**Цель работы**

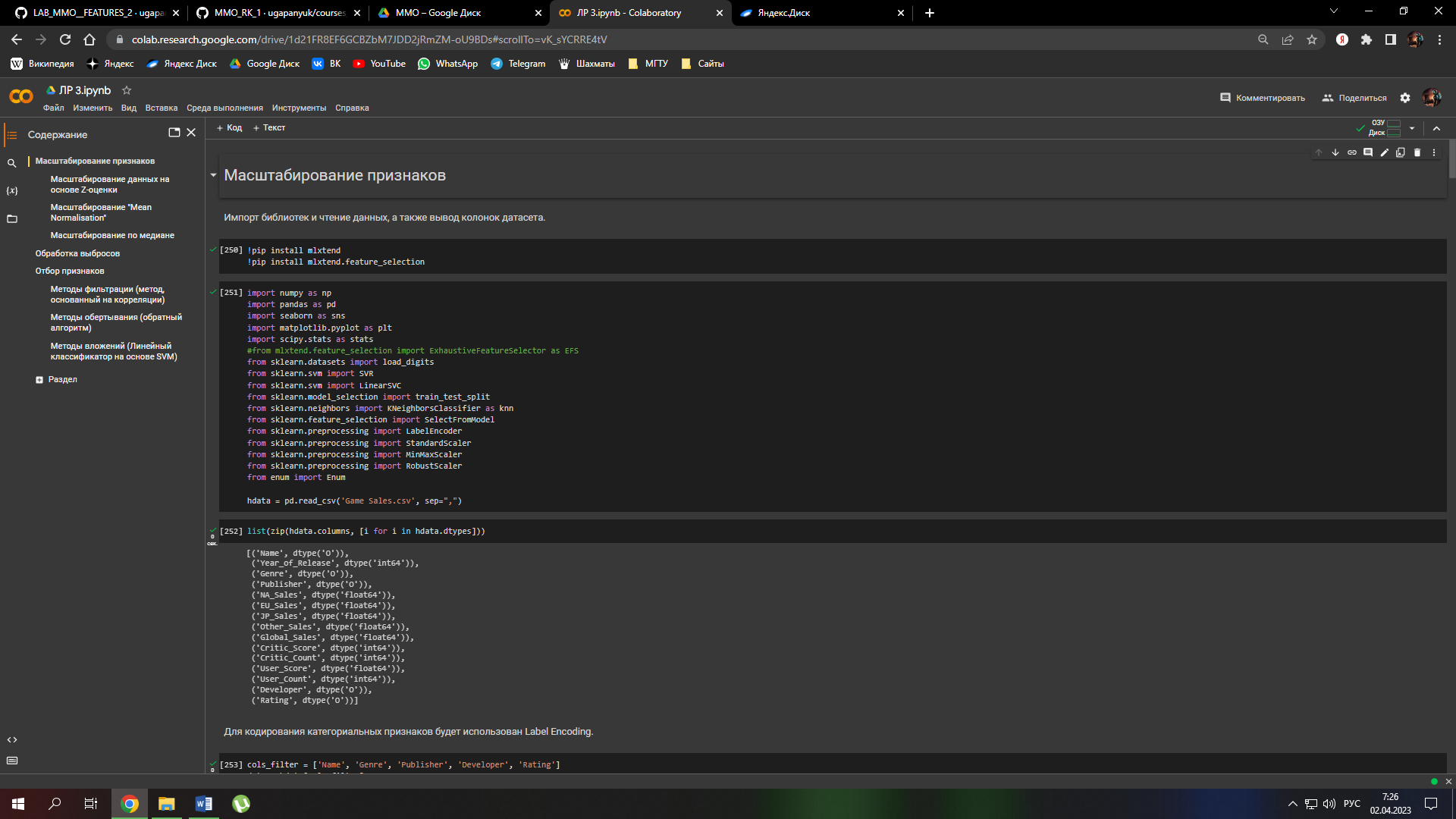
Изучение продвинутых способов предварительной обработки данных для дальнейшего формирования моделей.

**Задание**

1. Выбрать один или несколько наборов данных (датасетов) для решения следующих задач. Каждая задача может быть решена на отдельном датасете, или несколько задач могут быть решены на одном датасете. Просьба не использовать датасет, на котором данная задача решалась в лекции.
2. Для выбранного датасета (датасетов) на основе материалов лекций решить следующие задачи:

* масштабирование признаков (не менее чем тремя способами);
* обработку выбросов для числовых признаков (по одному способу для удаления выбросов и для замены выбросов);
* обработку по крайней мере одного нестандартного признака (который не является числовым или категориальным);
* отбор признаков:
* один метод из группы методов фильтрации (filter methods);
* один метод из группы методов обертывания (wrapper methods);
* один метод из группы методов вложений (embedded methods).

**Описание выполнения**



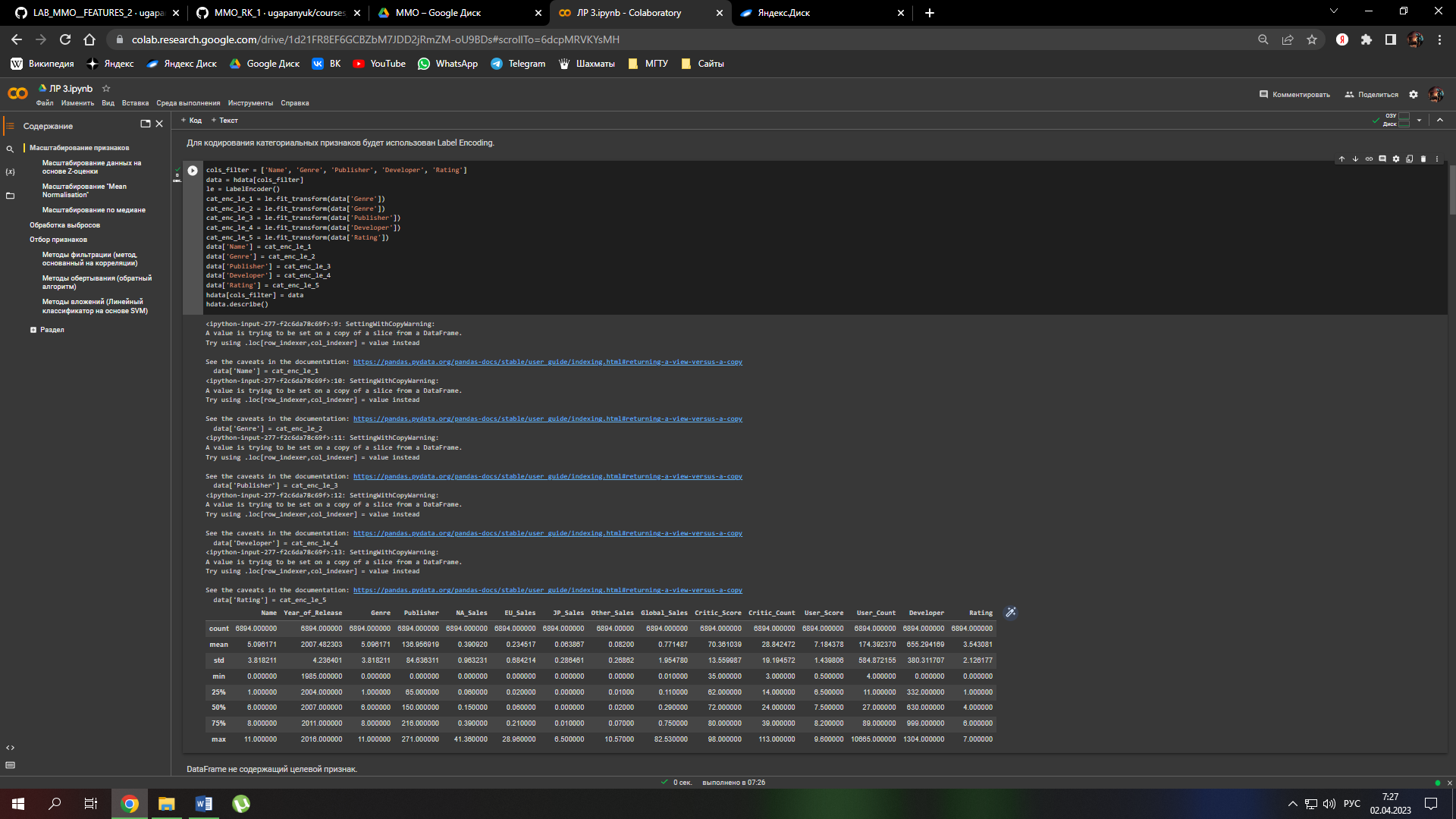


Рисунок 1-2 – Импорт данных и кодирование категориальных признаков

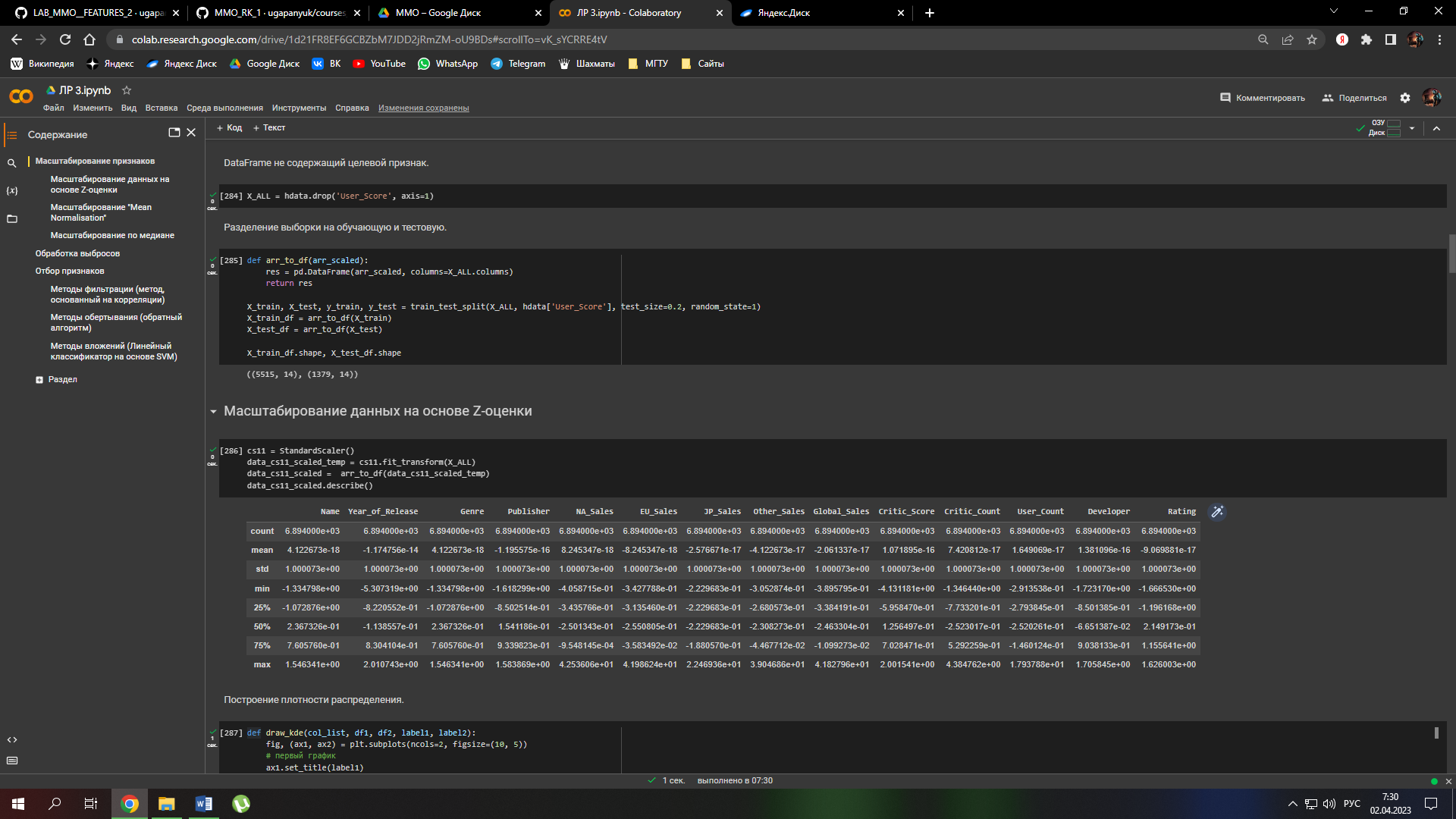
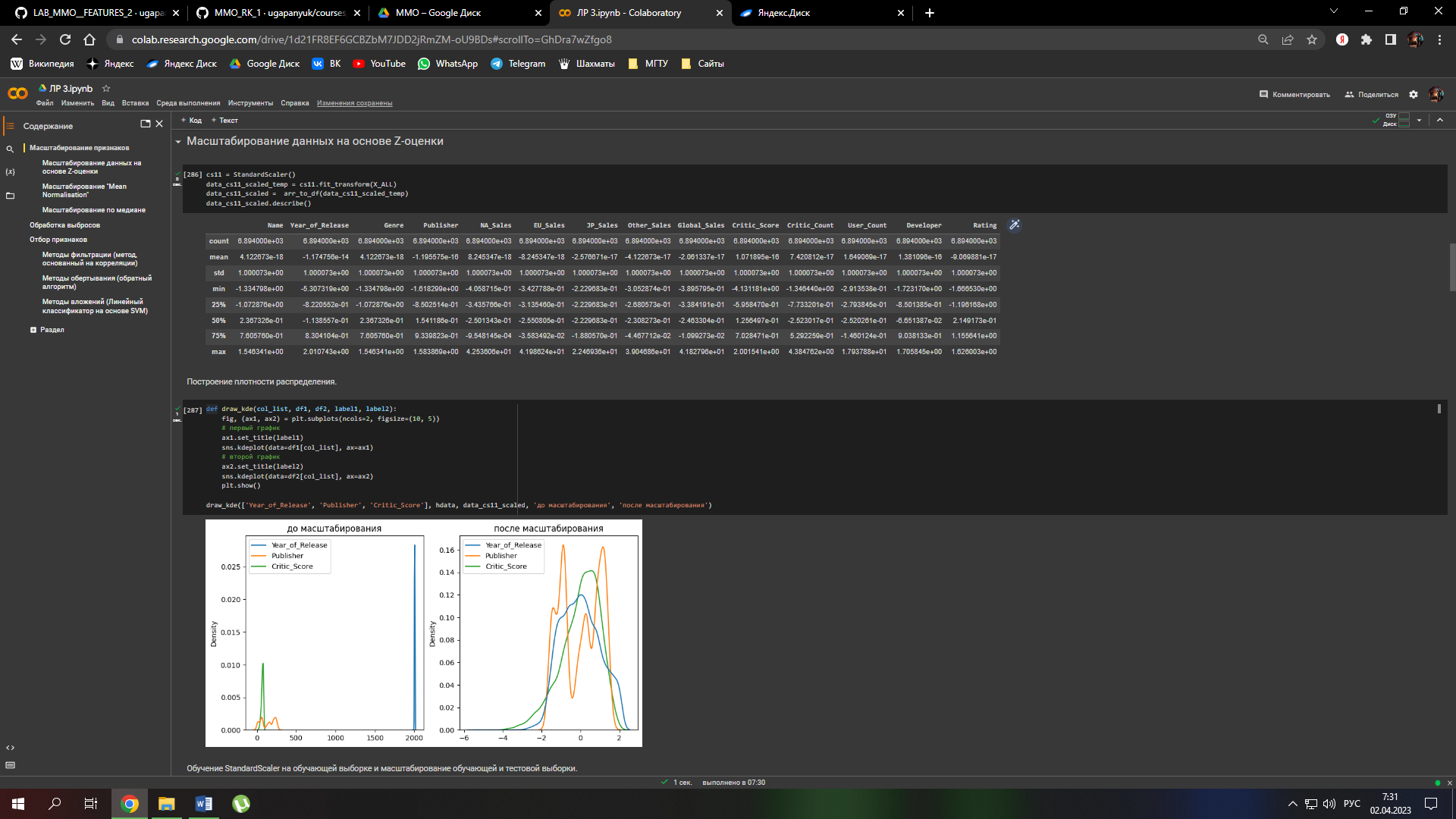


Рисунок 3 – Разделение выборки на обучающую и тестовую



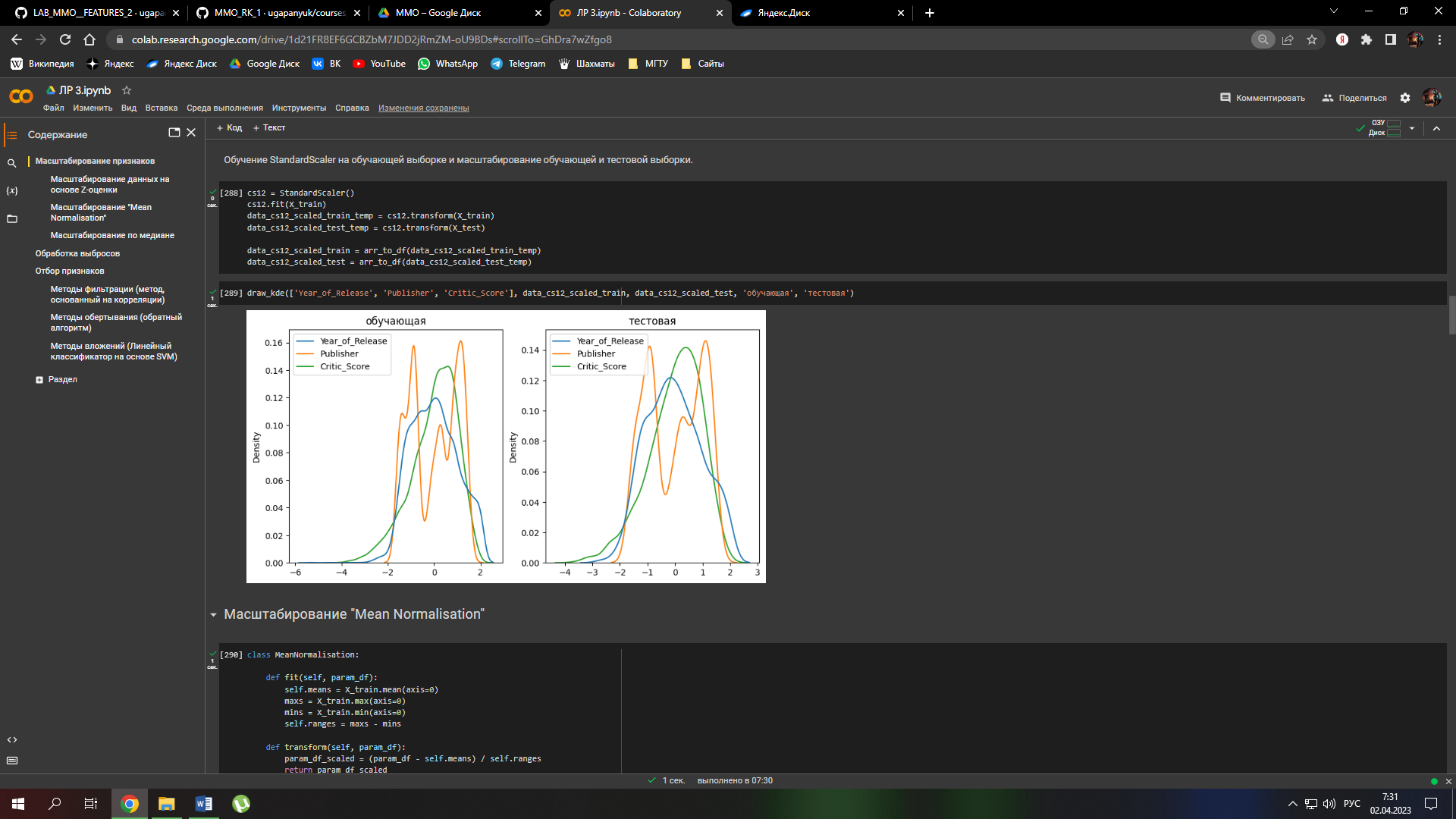


Рисунок 4-5 – Масштабирование данных на основе Z-оценки

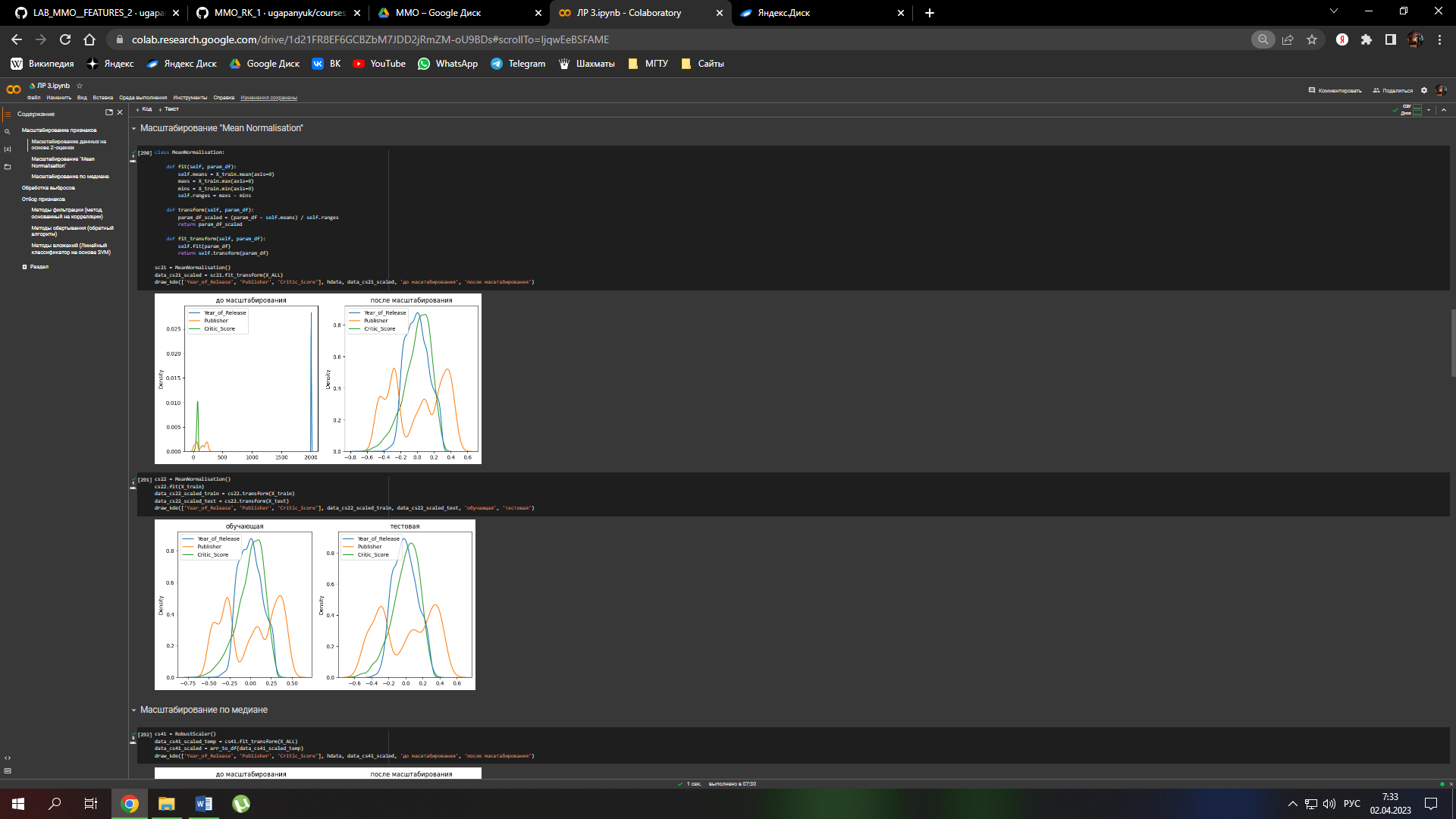


Рисунок 6 – Масштабирование "Mean Normalisation"

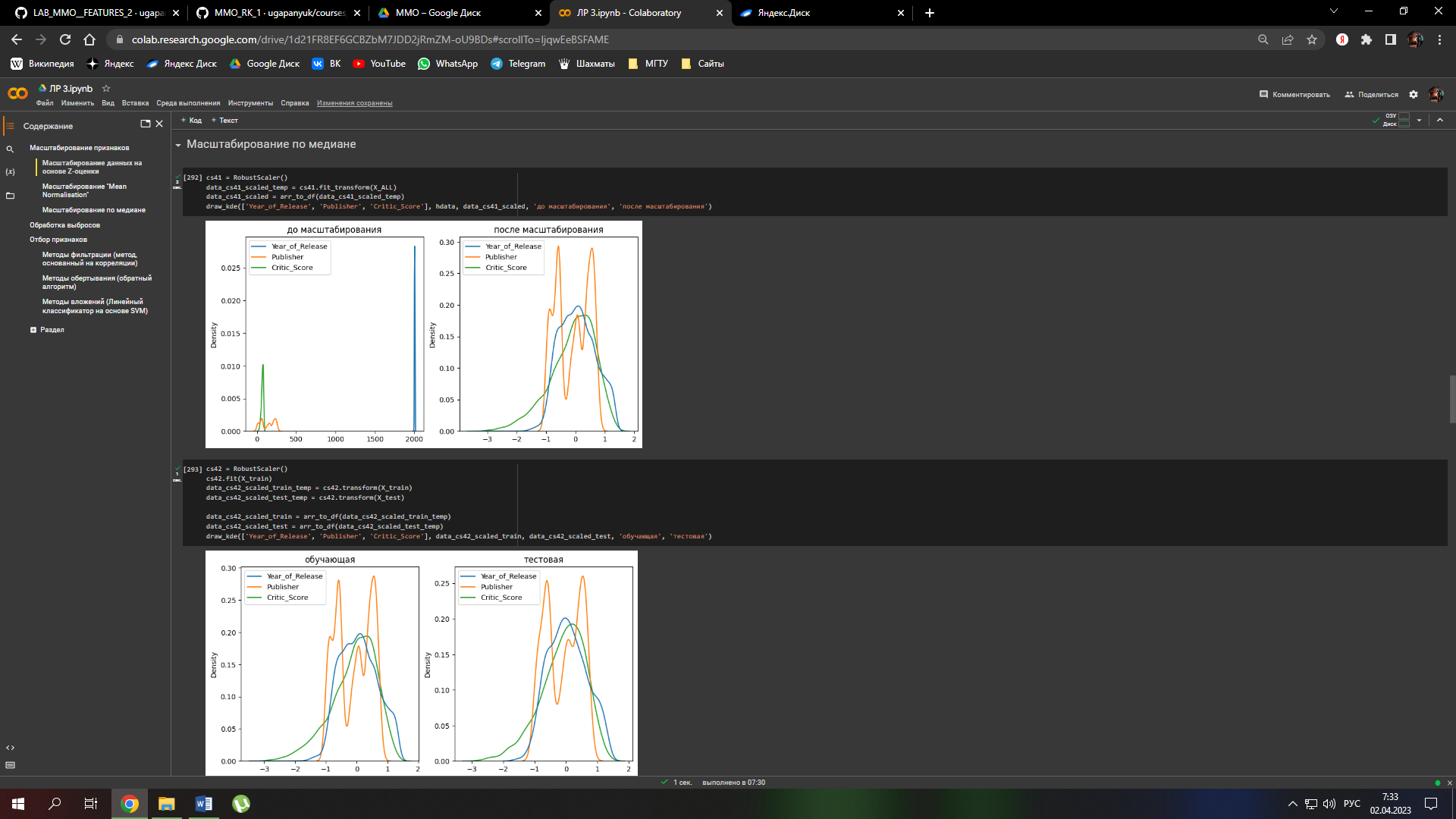
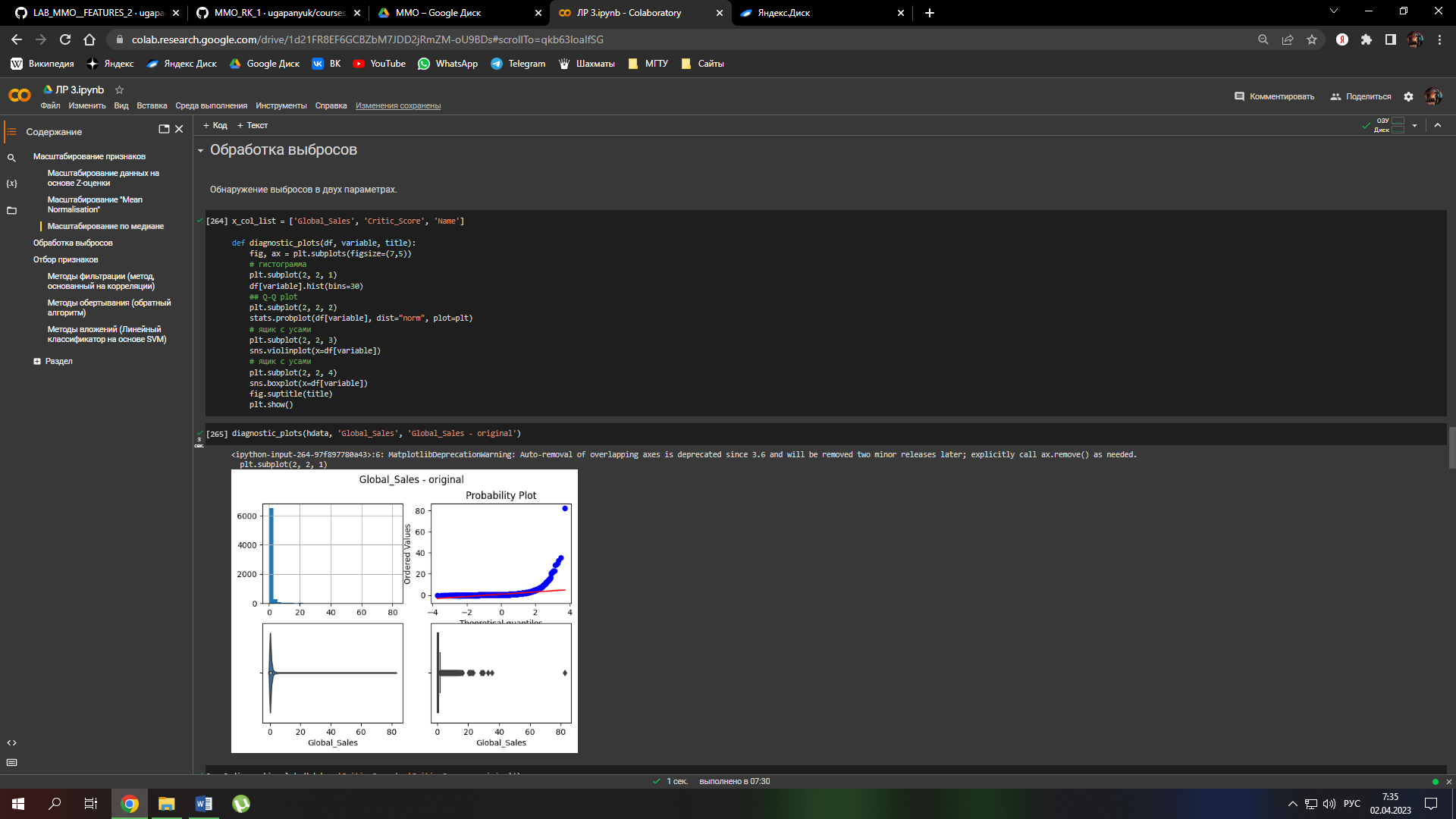


Рисунок 7 – Масштабирование по медиане



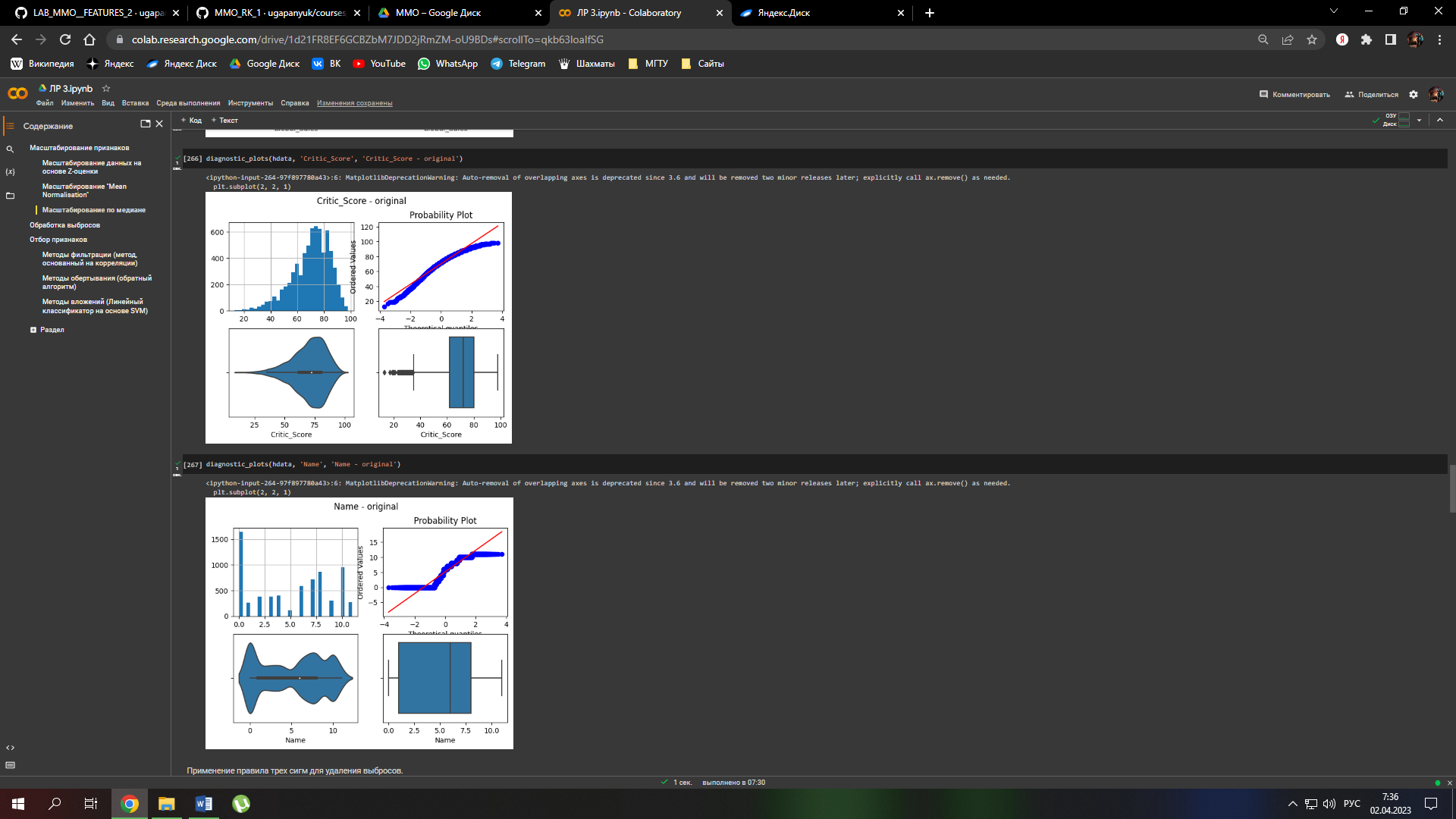
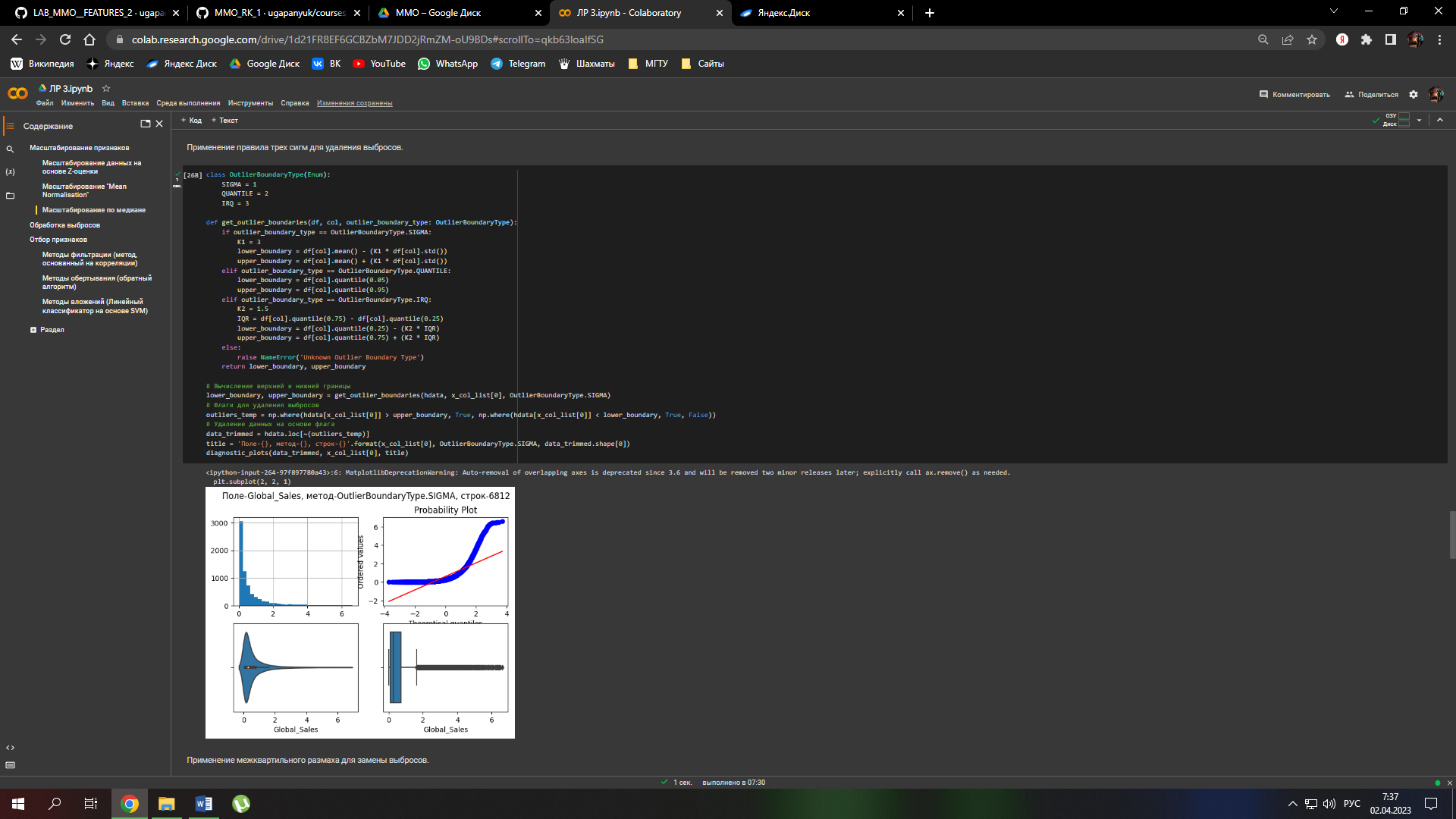


Рисунок 8-9 – Обнаружение выбросов в двух параметрах



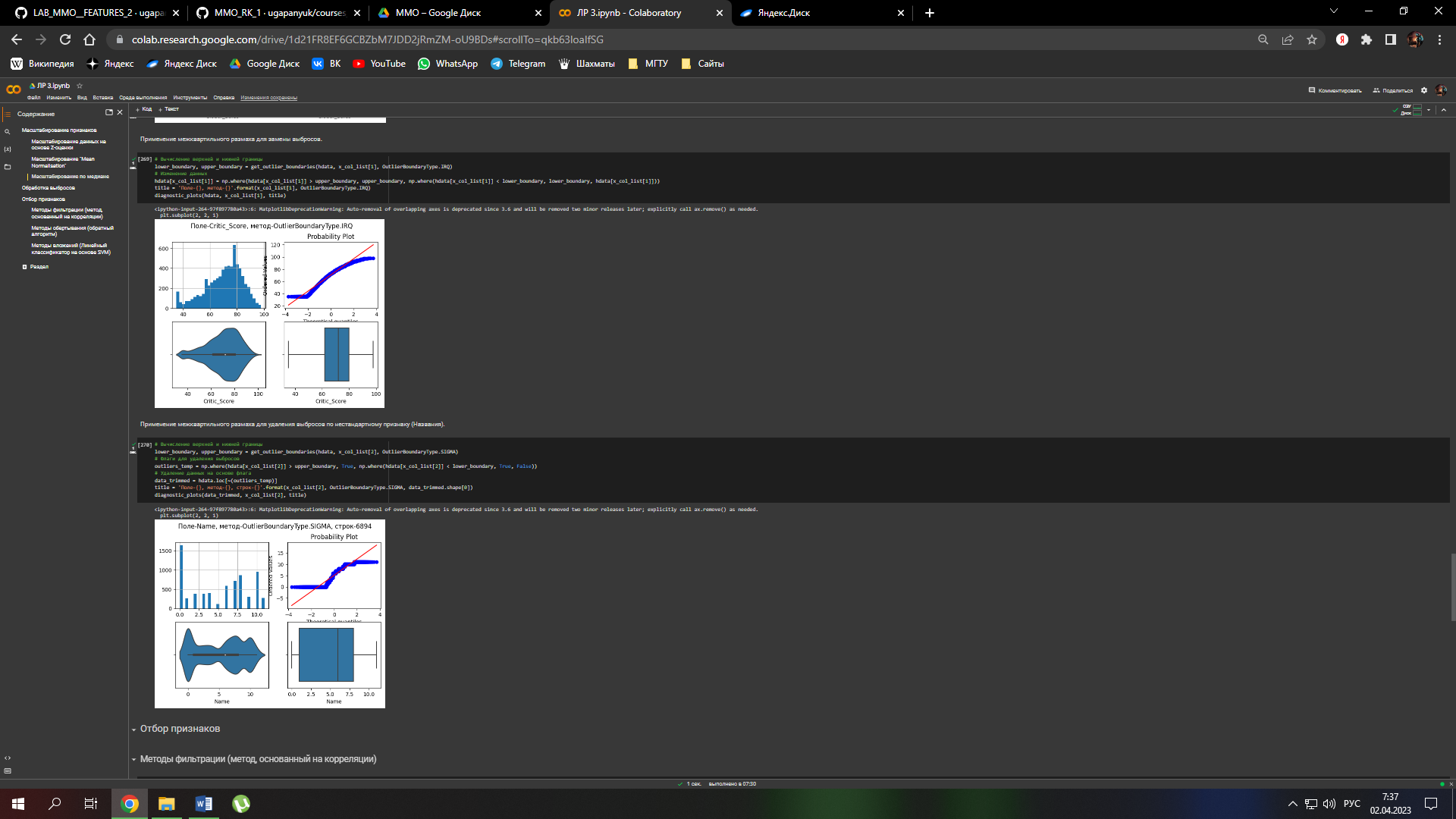
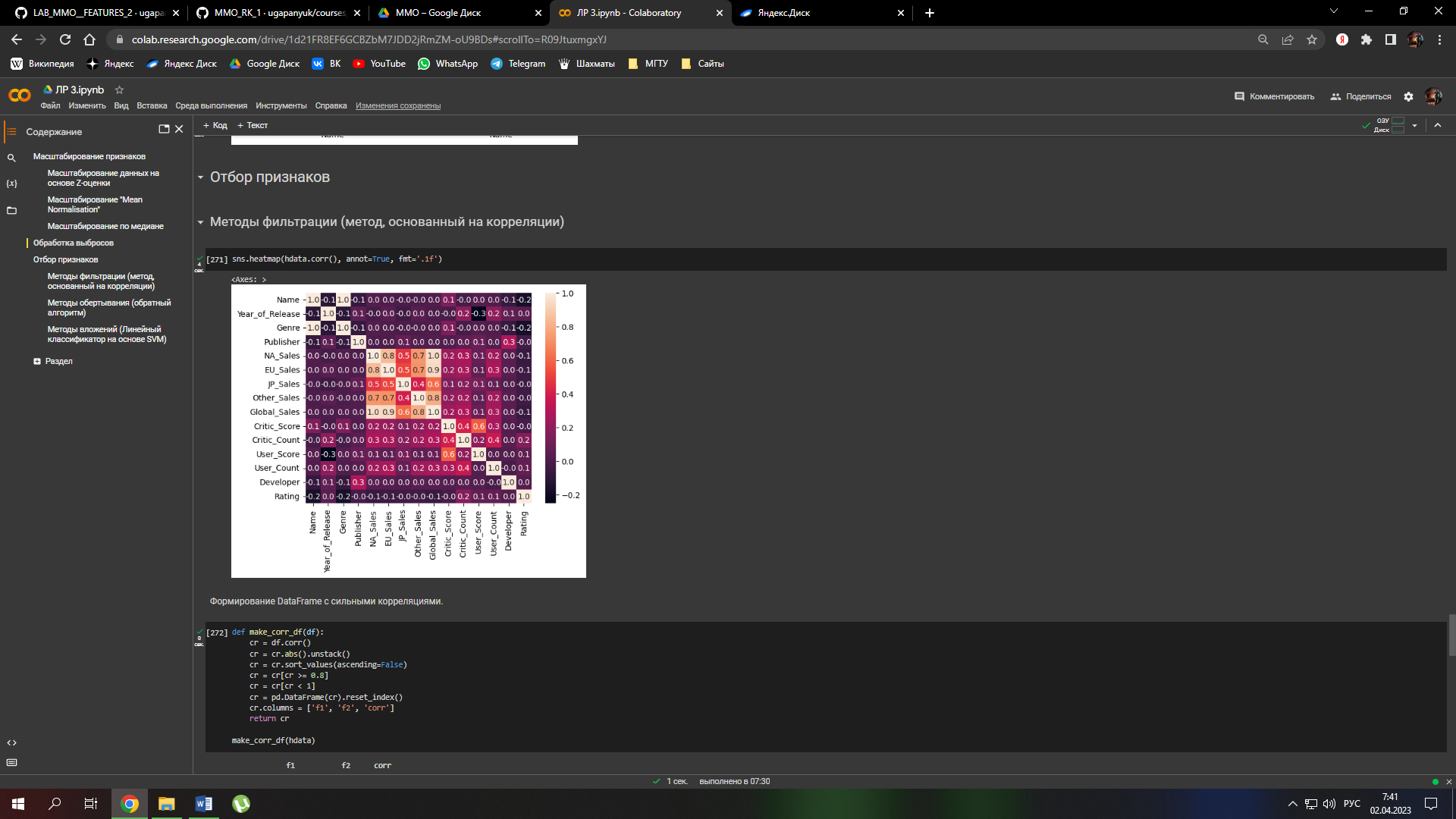


Рисунок 10-11 – Обработка выбросов



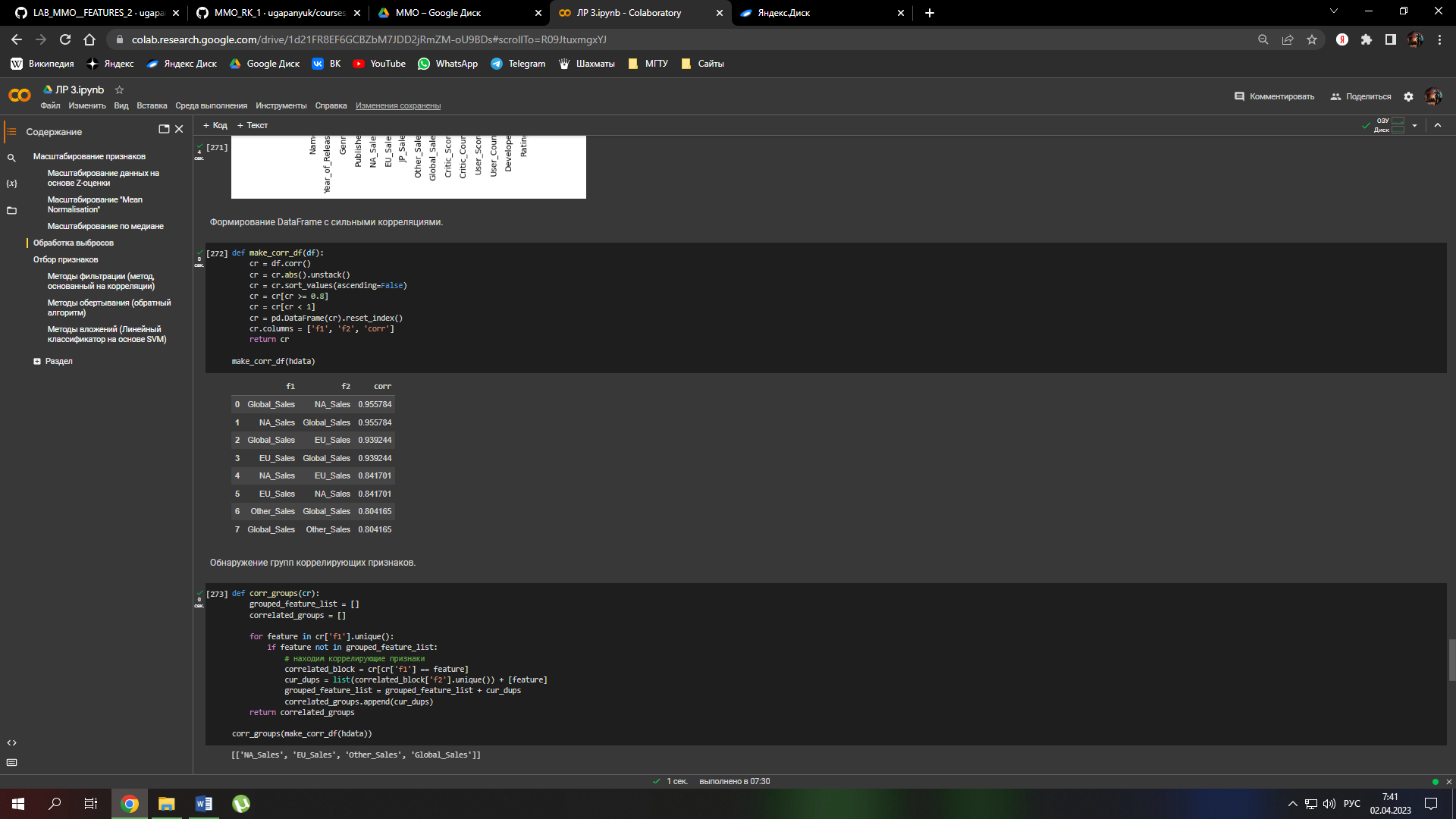


Рисунок 12-13 – Методы фильтрации

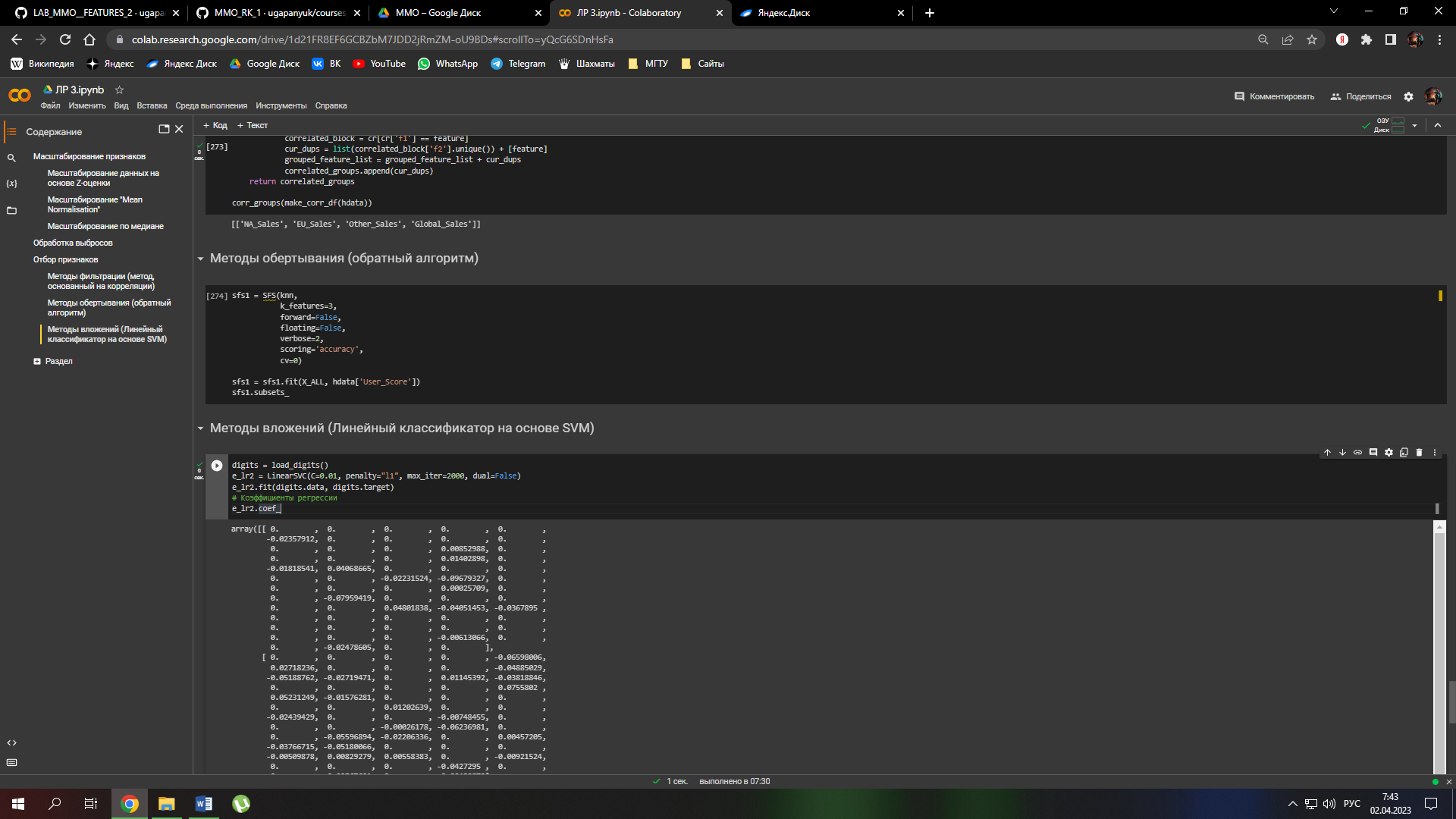


Рисунок 14 – Методы обертывания

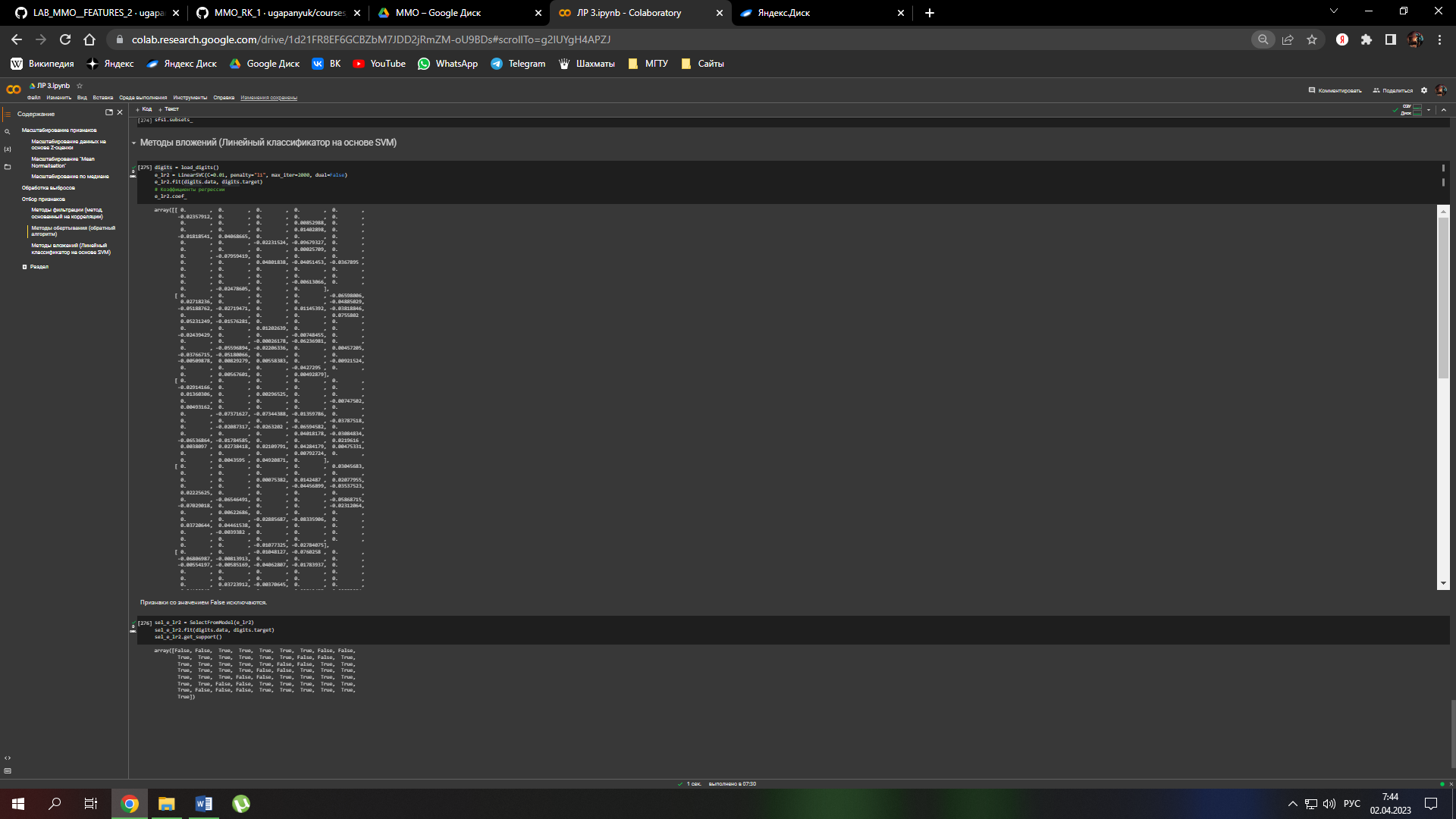


Рисунок 15 – Методы вложений

**Вывод**

Таким образом, удалось изучить продвинутые способы предварительной обработки данных для дальнейшего формирования моделей.